Project Presentation & Report Guidelines

STAT 672-001: Statistical Learning & Data Analytics

George Mason University, Spring 2018
Credit Hours: 3.00

Project Report Guidelines

Let me here repeat and further elaborate on what I mentioned in class and in the syllabus.

Scientific Writing.

The formal level of your writing should at least come close to that of a research or review article in an international journal. In particular, this entails the following:

(i) appropriate use of language and terminology,

(ii) precision and conciseness,

(iii) clear notation,

(iv) proper use of citations.

Regarding item (ii), your exposition should leave no room for interpretation/speculation. It is bad style to have the reader figure out what you actually mean. Conciseness implies that you avoid repetitions or other forms of redundancies, and that your exposition cannot be easily reduced without loss of information. Regarding notation, confine yourself to a minimum. Do not introduce notation that it is never referred to later on. Avoid ambiguities and inconsistencies. In particular, do not use multiple notations for the same object, nor use identical notation for different objects. Depending on your level of experience, it may be wise to stick to the notation employed in the underlying material. As a general rule, all external resources used to prepare your report need to be cited.

Scope and recommended structure.

The report typically starts with an introduction containing background and motivation, including the use of small illustratory (data) examples at your discretion. Avoid going too much into technical details at this point. Optionally, you can add an abstract at the very beginning, or a small separate paragraph on notation (standard notation as or very similar to the one used in class does not need to be explained). The introductory section is followed by a method/theory section and a data analysis/case study section. The final section contains a summary and a discussion (alternative directions, open problems, ...).

An overview of references follows at the very end. Technical details or supplemental data analyses can be placed into an appendix. Depending on the topic, I expect the report to have a length between 7-8 and 15 pages, assuming standard formatting and moderate use and size of figures/tables; going beyond 30 pages is strongly discouraged. These figures are only for your orientation. As mentioned in class, the quality, not the length, determine the grade. Very importantly, keep the scope comparatively narrow and avoid digressions into other topics, unless you think that this adds substantial value.

Do your own work.

The project report is not intended to be an exercise in reproducing existing material, nor in cleverly rephrasing existing material. The report should rather contain your own account of the topic you have selected. For this reason, it is advisable to compile your report with sufficient distance to the material you used for preparation. One recommendation is to identify sub-topics that you consider particularly interesting, and discuss these in more depth. You are welcome to address weak points or open problems in existing methods, but criticism needs to be carefully justified.

Do not replicate simulations or data analysis of the material unless you think that they serve an important illustratory purpose. Try to think of settings/data sets not already considered in the paper that may serve as additional illustration of what you want to convey about the topic.

Identify applications.

In your report, try to address the relevance of the approach for applications. For example, what has led to its development, or in what respect does it improve over more conventional tools.

Make connections to central themes of the class.

Here is a (incomplete) list of central themes: curse of dimensionality and phenomena in high dimensions, computational hardness and approximations, bias-variance trade-off, empirical risk minimization, regularization, sparsity, smoothness, parametric vs. non-parametric approaches, statistical robustness, dimensionality reduction, model selection and assessment. Clearly, it is your job to identify those which have a close link to your topic. Typically, it will not be more than one or two. If you think that there are several ones, please feel free to make a selection (it is completely acceptable to discuss a single theme). Ideally, you make the connections "along the way" rather than in a dedicated subsection.

Project Presentation Guidelines

The main goal of the presentation is to extract the very essence of your topic and explain it to your peers in such a way that they can connect it well to what they have encountered in class. Provide a nice take-home message.

In order to accomplish the above goal, you need to condense the material contained in the report considerably. When preparing the presentation, it is important to keep in mind that your peers know far less about the topic than yourself; you need to start from a common basis. Avoid technical details, but avoid being completely non-technical as well. I expect to see a few central and simple equations in every presentation. Be particularly mindful regarding notation.

It is expected that you will use between six and ten slides. Note the time limit of ten minutes. Because of that, you should thoroughly practice your presentation in advance (speaking it out loudly, alone or in the presence of listeners). Otherwise, you may lose valuable time thinking about what would be good to say during your talk. It is not acceptable to exceed your speaking time by a margin; you will be cut off at some point. The other extreme would be to memorize your talk word by word and rush over your slides.

Optimize the design of your slides (formatting, fonts, figures, etc.) with regard to readability.

Lastly, please show interest in what your peers present. Think about questions to ask at the end of their presentations. Unless it is something that can be answered quickly, do not interrupt the speaker.